Handoff Minimization Through a Relay Station Grouping Algorithm With Efficient Radio-Resource Scheduling Policies for IEEE 802.16j Multihop Relay Networks

Shun-Ren Yang, *Member, IEEE*, Chien-Chi Kao, *Student Member, IEEE*, Wai-Chi Kan, and Tzung-Chin Shih

***Abstract*—The IEEE 802.16j standard has been developed to provide performance enhancement to the existing IEEE 802.16e network by incorporating the multihop relay (MR) technology. However, frequent handoffs and low spectrum-utilization issues that were not encountered in IEEE 802.16e may be incurred in IEEE 802.16j. The relay station (RS) grouping is one optional mechanism in the IEEE 802.16j MR standard to overcome these problems. The concept of RS grouping is to group neighboring RSs together to form an RS group, which can be regarded as a logical RS with larger coverage. In this paper, we investigate RS grouping performance enhancement in terms of throughput and handoff frequency. This paper designs an RS grouping algorithm to minimize handoffs by utilizing a greedy grouping policy: RS pairs with higher handoff rates will have higher priority for selection. The simulation results show that the handoff frequency of the considered MR network can significantly be reduced, and suitable RS grouping patterns can be derived using our grouping algorithm. In addition, we propose two centralized scheduling policies, i.e., the throughput-first (TF) policy to maximize the system throughput and the delay-first (DF) policy to minimize the average packet delay. By integrating our RS grouping algorithm and centralized scheduling algorithms, the simulation results indicate that, for the case of fixed users, groupings with smaller group sizes can result in better throughput performance. However, when user mobility is considered, the throughput value increases as the group size increases. Furthermore, we also show that the DF policy can both minimize the average packet delay and provide the fairness property among users with different traffic loads.**

***Index Terms*—Grouping algorithm, IEEE 802.16j, multihop relay (MR), scheduling policy, WiMAX.**

Manuscript received March 19, 2009; revised August 18, 2009 and November 16, 2009; accepted January 08, 2010. Date of publication February 2, 2010; date of current version June 16, 2010. This work was supported in part by the National Science Council of Taiwan under Contract 96-2752-E-007-003-PAE, Contract 96-2221-E-007-025-, Contract 96-2221-E-007-027-, Contract 96-2219-E-007-012-, and Contract 96-2219-E-007-011-, and in part by Chunghwa Telecom. The review of this paper was coordinated by Prof. V. W. S. Wong.

S.-R. Yang is with the Department of Computer Science and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 300, Taiwan (e-mail: sryang@cs.nthu.edu.tw).

C.-C. Kao is with the Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan (e-mail: mickey@wmnet.cs.nthu.edu.tw).

W.-C. Kan is with Synology Inc., Taipei 10351, Taiwan (e-mail: waichik@gmail.com).

T.-C. Shih is with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 300, Taiwan (e-mail: stonc@wmnet.cs.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2010.2041677

I. Introduction

INCORPORATING the *multihop relay* (MR) technology [25], the IEEE 802.16j MR standard [14] has been developed to provide throughput improvement, coverage extension, and capacity enhancement to the existing IEEE 802.16e protocol [1]. By deploying relay stations (RSs), the end-to-end communication quality between base stations (BSs) and mobile stations (MSs) can be improved without high infrastructure deployment costs. In particular, it becomes possible to forward data to an MS using a high transmission rate in line-of-sight (LOS) conditions through an MR path to avoid the nonline-of-sight (NLOS) direct (i.e., single-hop) transmission with bad channel quality. In addition, *spatial reuse* [19] is another promising approach that can be employed in IEEE 802.16j MR networks to improve spectral efficiency. Based on the centralized scheduling, spatial diversity gain can be achieved if multiple simultaneous transmissions using the same bandwidth resources are realized within the same BS cell.

Although IEEE 802.16j has the potential to provide substantial performance enhancements, several issues that were not addressed in IEEE 802.16e may be encountered. For example, frequent handoffs may occur during the movements of MSs since the RS cells are smaller than the BS cells. Moreover, the unbalanced resource allocation of different RSs may result in inefficient spectrum utilization. To avoid the consequent performance degradation, *RS grouping* has been identified as an optional mechanism in the IEEE 802.16j standard. The main idea of RS grouping is to put adjacent RSs together to form an *RS group*, wherein the RS members are required to simultaneously receive and transmit the same data. From the MSs' viewpoint, the RS group can be regarded as a logical RS with larger coverage. Therefore, the handoff frequency can be reduced since no handoff procedure would be triggered, even when an RS crossing event within the RS group occurs. In addition, since the radio resources of the RS members are aggregated together, the resources can then be more flexibly and efficiently allocated to achieve higher spectrum utilization. However, to the best of our knowledge, no RS grouping strategy has been proposed or discussed in the literature. We argue that different grouping criteria may lead to various performance results. Specifically, utilizing a smaller RS group size is advantageous to spatial reuse, because more RS groups can perform simultaneous transmissions at the same time and frequency. Thus, improved average

0018-9545/$26.00 © 2010 IEEE

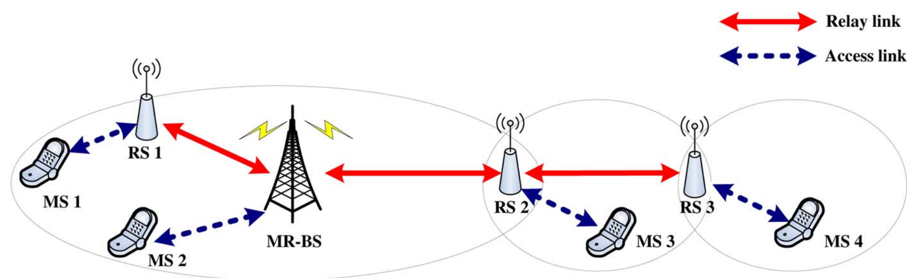

Fig. 1. Sample topology of IEEE 802.16j MR networks.

system throughput can be expected. However, a smaller RS group size may also lead to higher packet loss rate due to more frequent handoffs between RS groups. In conclusion, when implementing the RS grouping mechanism in IEEE 802.16j MR networks, the performance tradeoff between throughput and handoff frequency should seriously be considered.

In this paper, we analyze grouping strategies for RS-grouping-enabled IEEE 802.16j MR networks and propose an efficient RS grouping algorithm to minimize the handoff frequency. As we have pointed out, RS grouping strategies will also influence the throughput performance. To investigate the impacts of RS grouping on the IEEE 802.16j system throughput, we design two centralized downlink (DL) scheduling policies for RS-grouping-enabled IEEE 802.16j MR networks. One of these two scheduling policies aims to maximize the system throughput, and the other is to minimize the average packet delay. The throughput estimation results under different grouping configurations can assist network service providers to choose the most appropriate settings of grouping factors (e.g., group size). Notice that the spatial reuse concept is considered during the scheduling procedure to improve spectral efficiency. By integrating our proposed RS grouping algorithm and centralized scheduling policies, the simulation results show that the throughput and delay performance can be improved in addition to the significantly reduced handoff frequency. The main contribution of this paper is to propose the first integrated algorithmic framework that can be utilized to investigate the performance interaction between RS grouping and resource scheduling for IEEE 802.16j MR networks.

The remaining parts of this paper are organized as follows: In Sections II and III, we give an overview of the IEEE 802.16j standard and a discussion of the IEEE 802.16j RS grouping mechanism, respectively. Section IV presents an RS grouping strategy analysis and details our proposed RS grouping algorithm. In Section V, two efficient centralized DL scheduling policies for IEEE 802.16j MR networks are addressed. The performance evaluations of our grouping algorithm and scheduling policies are presented in Section VI. Section VII summarizes the related work. Finally, we conclude this paper in Section VIII.

II. IEEE 802.16j Multihop Relay Networks

A. Network Architecture

The IEEE 802.16j standard is expected to enhance the system performance of IEEE 802.16-based networks through multihop-relaying technologies. A typical topology example of IEEE 802.16j MR networks is shown in Fig. 1. In this network, an MS can access the MR-BS either through a multihop-relaying path (e.g., MS1, MS3, and MS4) or directly (e.g., MS2). In addition, a station (BS or RS) is called an *access station* if it provides network attachment functionality to a given MS or RS. On the other hand, an RS is a *subordinate* RS of another station if that station serves as the access station for that RS. For instance, RS2 is the access station of RS3, and RS3 is a subordinate RS of RS2. The wireless links that directly connect access stations with their respective subordinate RSs are called *relay links*, whereas the links between MSs and their corresponding access stations are known as *access links*. Since an RS can only be subordinated to one station, the MR-BS and the RSs in this MR network form a tree-based MR topology. Note that it has been shown that the throughput decreases as the number of hop counts increases [7]; thus, we only investigate two-hop IEEE 802.16j networks in this paper.

B. Frame Structure

The frame structure of IEEE 802.16j MR systems is extended from that of IEEE 802.16e networks, which also adopt orthogonal frequency-division multiple access (OFDMA) as the primary channel access mechanism for NLOS communication. The basic unit of resource for allocation in OFDMA is a *slot*, which comprises a number of symbols in the time domain and one subchannel in the frequency domain. The timeline is divided into contiguous frames, each of which further consists of a DL and an uplink (UL) subframes. In IEEE 802.16j, the DL and UL subframes shall include one *access zone* for MR-BS ↔ RS and MR-BS ↔ MS transmissions and may include one *relay zone* for RS ↔ subordinate MS transmissions, respectively.

III. IEEE 802.16j Relay Station Grouping Mechanism

Although deploying RSs in IEEE 802.16 networks can provide significant throughput or coverage enhancements, several issues regarding the relaying architecture of IEEE 802.16j should be addressed. These issues include frequent handoffs, redundant control overhead, and low spectral efficiency. It is perceived that these issues will result in unpredictable performance reduction for IEEE 802.16j MR networks. Therefore, the IEEE 802.16j standard provides the optional RS grouping mechanism to reduce the impacts of these issues. The concept of an RS grouping is that adjacent RSs could be grouped together as an RS group, which acts as a virtually regular RS to its associated MSs. The grouping criteria are decided by the controlling MR-BS, based on the targeted performance

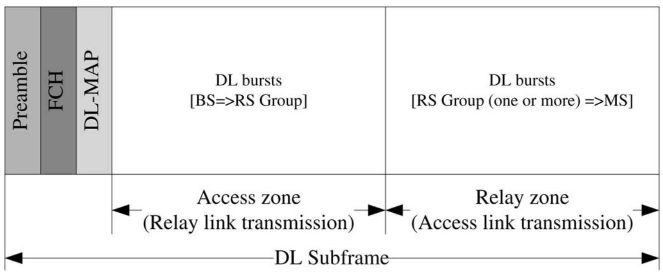

Fig. 2. Modified IEEE 802.16j DL subframe structure for supporting RS grouping.

requirement. Note that the coverage of an RS group is larger than that of its regular member RSs, and no handoff event would be triggered, even though an RS-cell crossing event within the same RS group occurs. Consequently, the MSs under the RS grouping mechanism will experience lower handoff probability. On the other hand, MR-BSs can manage RS groups using only one set of control header; hence, the control signal overheads are reduced. Finally, the radio resource (i.e., the relay zone) of each member RS can be aggregated and shared by all the MSs under the corresponding RS group so that the spectral efficiency is improved.

For the DL operation, the member RSs of a group should be configured to transmit equivalent data signals to the same MS. Thus, the subordinated MSs can receive the best quality signal within the group, no matter where they are located. This operation is also called cooperative transmission, because the member RSs will form a virtual antenna array to exploit macrodiversity. On the other hand, the diversity combining of the information received by the member RSs of an RS group can be performed in the UL situation. Both the DL and UL diversity gains can be achieved under a centralized scheduling scheme by keeping the MS list of each RS group at the respective MR-BS. Therefore, the RS grouping mechanism is reasonable in improving the data-transmission rate.

To support an RS grouping, the original IEEE 802.16j frame structure should be modified (see Fig. 2 for the modified DL subframe). Specifically, access zones should handle the transmissions between BSs and RS groups, whereas relay zones should handle the transmissions between RS groups and subordinate MSs. From the mobility management point of view, we note that implementing RS grouping will not incur extra costs to IEEE 802.16j. As previously mentioned, for movement between the RS cells of an RS group, an MS will not initiate the handoff procedure. The MS CDMA periodic ranging process with aggregated ranging subchannel allocation [13] can be employed to handle the RS reselection during intra-RS-group movement. On the other hand, when the MS roams from an RS group to another RS group, since an RS group can be seen as a legacy BS, the conventional MAC layer handoff procedure is directly applied for this scenario.

IV. Our Proposed Relay Station Grouping Algorithm for IEEE 802.16j Multihop Relay Networks

This section proposes an IEEE 802.16j RS grouping algorithm to reduce the handoff frequency of mobile users under prescribed IEEE 802.16j MR network performance requirements. The system assumptions and the concept of our proposed algorithm are described in the first section. Then, the proposed algorithm is discussed as a three-phase procedure in the succeeding sections.

A. System Assumptions and the Concept of Our RS Grouping Algorithm

To accommodate general scenarios, our algorithm makes no assumptions of the underlying IEEE 802.16j MR network topology, the user mobility behavior, and/or the packet traffic pattern. Specifically, within a considered BS, the IEEE 802.16j RSs can arbitrarily be deployed, and the coverage area of each RS can be irregular. In addition, the MSs within the considered BS can randomly move. In such an arbitrary environment, we only require that the handoff-rate information between each two RS cells should be available. The handoff rate between two RS cells, i.e., RS cell 1 and RS cell 2, is the total rate that the resident MSs hand off from RS cell 1 to RS cell 2 or from RS cell 2 to RS cell 1. Note that the handoff-rate information can simply be derived from the statistical data that are collected by the network service providers.

To design the RS grouping algorithm in our considered environment, we first specify the factors that may affect the grouping result. First, the group size is a factor that has the potential to influence the spatial diversity gain and the handoff frequency. It can be observed that utilizing a smaller group size will result in more RS groups. Such a grouping policy is beneficial to spatial reuse, although it causes higher handoff frequency. On the other hand, a larger group size has reverse effects on the spatial reuse and the handoff frequency, respectively. In addition to group size, the selection order of group members is also significant to the grouping strategy design. Even with the same group size, different grouping orders may lead to different numbers of handoff events. Once an RS grouping layout is provided by applying the most desired group size and grouping order, determining which set of RS groups to simultaneously transmit data further critically impacts the spectrum efficiency. In this paper, we define an *activation set* as a particular set of RS groups, which can transmit data to their respective resident MSs at the same time and frequency without interference. Clearly, the assignment of activation sets is another important factor that should be considered in the design of the RS grouping algorithm.

Taking all the preceding factors into account, we propose our RS grouping algorithm, which contains three phases, as shown in Fig. 3: 1) *grouping phase*; 2) *activation-set-assignment phase*; and 3) *scheduling-simulation phase*. Based on the handoff-rate information, the grouping phase first constructs the handoff-minimizing RS grouping result for a preferred group size. Then, the grouping phase generates the adjacency matrix of the constructed RS groups. Based on the adjacency matrix, the activation-set-assignment phase assigns each RS group to an activation set to enhance the spatial diversity gain. Afterward, in the scheduling-simulation phase, the DL transmission simulation is executed for the given grouping result with determined activation-set assignment. By comparing the performance gains

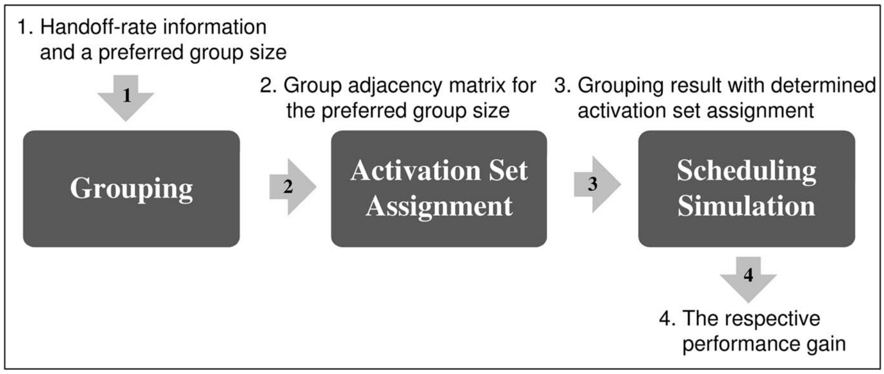

Fig. 3. Concept of the proposed RS grouping algorithm.

Algorithm 1 THE PROPOSED RS GROUPING ALGORITHM

```
1: Input:
2: α (the set of RSs within a considered BS), H (the handoff-rate matrix of the RSs in α).
3:
4: Output:
5: Ω (the desired grouping result with determined activation set assignment).
6:
7: Initialization:
8: Ω ← NIL;
9: key ← 0;
10: M ← rows[H];
11:
12: Procedure:
13: for i = 1 to M do
14:     Γ_i, A_i ← Grouping(α, H, i);
15:     Ω_i ← ActivationSet(A_i);
16:     S_i ← Scheduling(Γ_i, Ω_i);
17:     if ω(S_i) > key then
18:         key ← ω(S_i);
19:         Ω ← Ω_i;
20:     end if
21: end for
```

Algorithm 2 $Grouping(\alpha, H, X)$

```
1: Input:
2: α (the set of RSs within a considered BS), H (the handoff-rate matrix of the RSs in α), X (the preferred group size).
3:
4: Output:
5: Γ (a set of constructed RS groups {g_1, g_2, ..., g_n, ...}), A (the adjacency matrix of the constructed RS groups in Γ).
6:
7: Initialization:
8: Temp_α ← α;
9: M ← rows[H];
10: n ← 0;
11:
12: Procedure:
13: /* Each RS group is constructed based on the handoff-rate matrix H. */
14: while there exist pairs of ungrouped adjacent RSs in Temp_α do
15:     n ← n + 1;
16:     Find the ungrouped adjacent RS pair (a, b) that has the highest handoff rate H[a][b];
17:     if a.TotHR(Temp_α) ≥ b.TotHR(Temp_α) then
18:         Remove a from Temp_α and add it into the nth group g_n;
19:     else
20:         Remove b from Temp_α and add it into the nth group g_n;
21:     end if
22:     while |g_n| < X and group g_n has ungrouped neighboring RSs do
23:         Find the neighboring RS c with the highest total handoff rate to/from group g_n;
24:         Remove c from Temp_α and add it into g_n;
25:     end while
26:     Add g_n into Γ;
27: end while
28: /* An isolated RS forms an RS group itself. */
29: while there exists an isolated ungrouped RS d in Temp_α do
30:     n ← n + 1;
31:     Remove d from Temp_α and add it into the nth group g_n;
32:     Add g_n into Γ;
33: end while
34: /* Initialize the RS-group adjacency matrix A. */
35: for i = 1 to n do
36:     for j = 1 to n do
37:         A[i][j] ← 0;
38:     end for
39: end for
40: /* Compute the RS-group adjacency matrix A. */
41: for k = 1 to M do
42:     for l = 1 to M do
43:         if the handoff rate H[k][l] > 0, RS k ∈ g_i and RS l ∈ g_j then
44:             A[i][j] ← 1;
45:         end if
46:     end for
47: end for
```

of the scheduling results from different preferred group sizes, network service providers can finally choose the most favorable grouping result with the associated activation set assignment as their IEEE 802.16j MR network configuration.

The operation of our RS grouping algorithm is depicted in Algorithm 1, where two input parameters, i.e., the set α of all the RSs within a considered BS and the handoff-rate matrix H of the RSs in α, should first be provided. In H, the entry $H[i][j]$ represents the total handoff rate between RS cell i and RS cell j. After variable initialization, the three algorithm phases at lines 14–16 are repeated for each preferred group size. Then, at line 17, the weighted function $\omega(S_i)$ calculates the weighted performance gain for a simulation result S_i. The highest performance gain is recorded in key at line 18, and the best grouping result with determined activation set assignment is stored in Ω at line 19. When Algorithm 1 terminates, the desired grouping result can thus be obtained. In the succeeding sections, we individually present the details of each phase.

B. Grouping Phase

The operation of the grouping phase is shown in Algorithm 2 and is explained as follows: According to the functionalities, the RS grouping procedure is partitioned into four main portions.

1) *Input parameters:* At the beginning of the procedure (line 2), three parameters α, H, and X are required as the inputs, where X denotes the preferred group size.
2) *Output results:* The outputs of the procedure (line 5) are the final grouping result $\Gamma = \{g_1, g_2, \ldots, g_n, \ldots\}$ and the adjacency matrix A of the constructed RS groups, where g_n denotes the nth RS group.

3) *Initialization stage:* In this stage (lines 8–10), some variables must be initialized before executing the grouping procedure. We use $Temp_\alpha$ to denote the set of ungrouped RSs within the considered BS. $Temp_\alpha$ is initialized as α. Then, we use M to denote the total number of RSs in α. Moreover, the variable n is initialized as 0.
4) *Procedure:* After initialization, the main grouping procedure is started from line 13 to line 47. Since the purpose of our algorithm is to minimize the handoff probability, we introduce a greedy grouping policy, under which RS pairs with higher handoff rates (provided from H) will have higher priority to be selected. A detailed description of the steps is given as follows:
   a) Group-construction loop for nonsingle-RS groups (lines 14–27): This loop constructs RS groups from the ungrouped adjacent RS pairs. The ungrouped adjacent RS pair with the highest handoff rate is considered first (line 16). Of this pair, the RS that has a higher total handoff rate to/from other RSs in $Temp_\alpha$ is selected to be the starting member of the nth group g_n (lines 17–21). Then, the ungrouped neighboring RSs of g_n are selected to join g_n in descending order of the total handoff rate to/from g_n. This join procedure (lines 22–25) is repeated until the group size of g_n is equal to the preferred group size X or g_n has no ungrouped neighboring RS. This group-construction loop is iterated to form RS groups until no pair of ungrouped adjacent RSs exists.
   b) Group-construction loop for isolated RSs (lines 29–33): This loop constructs RS groups for isolated ungrouped RSs. Each isolated ungrouped RS forms an RS group itself (line 31).
   c) Group adjacency matrix initialization (lines 35–39): Each entry of the group adjacency matrix A is initialized as 0.
   d) Group adjacency matrix computation (lines 41–47): After initialization, the adjacency matrix A is computed by examining whether a member of a group g_i is adjacent to a member of another group g_j (line 43). If yes, the entry $A[i][j]$ is set to 1 (line 44). After examining all the M RSs, the adjacency matrix A of the RS groups in Γ can be derived.

After the preceding grouping procedure, the grouping result Γ and the group adjacency matrix A can be obtained for use in the subsequent activation-set-assignment phase. We note that the time complexity of the grouping phase is mainly dominated by the two nested loops in lines 14–27 and lines 41–47. For the nested loop in lines 14–27, the numbers of iterations for both the outer and inner loops are no larger than $|\alpha|$. On the other hand, for the nested loop in lines 41–47, the numbers of iterations for the outer and inner loops are both equal to $|\alpha|$. Therefore, the overall time complexity of the grouping phase is $O(|\alpha|^2)$.

C. Activation-Set-Assignment Phase

Before elaborating on how our activation-set-assignment phase can generate appropriate assignment results, we first point out that the number of activation sets significantly affects the data-transmission performance. Specifically, the fewer the activation sets, the more RS groups there will be that can simultaneously transmit data. Consider the RS-group layout in Fig. 4(a), where its RS-group adjacency graph is shown in Fig. 4(b). Clearly, the assignment $\{\{g_1, g_3, g_5\}, \{g_2, g_4, g_6\}\}$ is better than another assignment $\{\{g_1, g_4\}, \{g_2, g_5\}, \{g_3, g_6\}\}$ since the former assignment contains fewer activation sets and, thus, has improved spatial diversity gain.

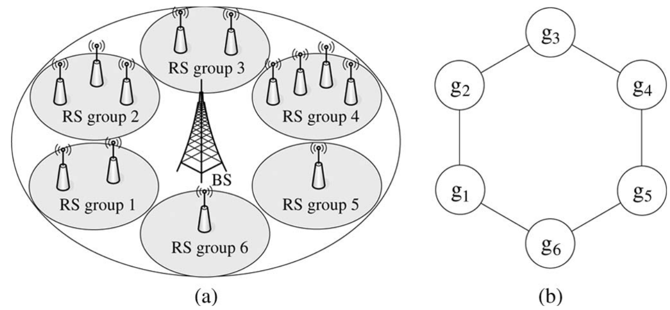

Fig. 4. Activation-set-assignment example.

Given an RS grouping result Γ and the corresponding adjacency matrix A from Algorithm 2, we first derive the adjacency graph $G(V, E)$, where V represents the set of RS groups, and E represents the interference relation among the RS groups. To minimize the number of activation sets, we model the activation-set-assignment problem as the minimum coloring problem of G. The rule to color all the vertices of G is that no two adjacent vertices can share the same color. Based on this rule, each set of the vertices that are assigned the same color is equivalent to an activation set. However, coloring G with the minimum number of colors is known as an NP-hard problem. To achieve a polynomial computation time, a well-known greedy coloring algorithm, i.e., the Welsh–Powell algorithm [6], is adopted in this paper. For the self-containedness purpose, we briefly summarize the key concept of the Welsh–Powell algorithm here. First, all possible colors are numbered. Then, following the descending order of the vertex degree, each vertex in G is sequentially selected to be assigned a color. When a vertex v attempts to be colored, the first color is examined to check if it has already been occupied by any of v's neighbors. If no, the first color is assigned to v. Otherwise, the next unoccupied color (by v's neighbors) is used to color v. This procedure is repeated for the subsequent vertices until the vertices in G have all been processed. The time complexity of the Welsh Powell algorithm is proven to be $O(|V|^2)$. This algorithm guarantees that the number of required colors is at most one more than the maximum degree $\Delta(G)$ of G. That is, the Welsh–Powell algorithm determines at most $\Delta(G) + 1$ activation sets. Our activation-set-assignment phase applies the Welsh–Powell algorithm and therefore can assign each constructed RS group (from the grouping phase) to an appropriate activation set.

D. Scheduling-Simulation Phase

In this section, we characterize the concept of scheduling simulation. By adopting a specific scheduling policy, scheduling simulations are conducted for the different grouping

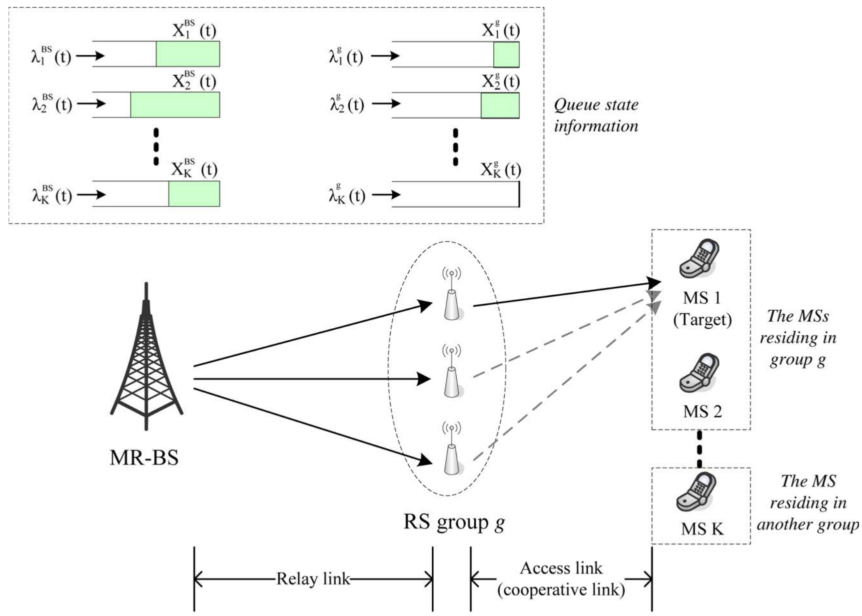

Fig. 5. System model of RS-grouping-enabled IEEE 802.16j MR networks.

results with associated activation set assignments provided from the grouping and activation-set-assignment phases. From these simulations, several output measures (e.g., system throughput and packet delay time) can be derived. The network service providers can evaluate the performance gains of these grouping results by assigning weighted values to the respective output measures following their preference rules. Then, by comparing the weighted indexes from the different grouping results, the most desirable grouping result can be determined. In the succeeding sections, the scheduling policies for RS-grouping-enabled IEEE 802.16j MR networks will be elaborated upon in Section V, and the representative simulation results will be presented in Section VI.

V. Multihop Centralized Downlink Scheduling Policies

In the scheduling-simulation phase of our RS grouping algorithm described in Section IV, centralized DL scheduling policies are required to evaluate the performance gain of a given RS grouping layout. In this section, we first define the scheduling problem for RS-grouping-enabled IEEE 802.16j MR networks. Then, the system description of our considered multihop network is addressed. Finally, we propose two centralized DL scheduling policies for IEEE 802.16j MR networks under RS grouping and spatial-reuse assumptions, with the objectives of maximizing the system throughput and minimizing the DL traffic delay, respectively.

A. Scheduling Problem for RS-Grouping-Enabled IEEE 802.16j MR Networks

Consider the DL transmission of an IEEE 802.16j MR network with a BS and a set of preconfigured RS groups, as shown in Fig. 5. Assume that a number of packets from the external networks are desired to be delivered to MS 1 residing in RS group g. The BS, which acts as the network gateway for its served MSs, will first buffer these packets for MS 1 in the corresponding packet queue. In an appropriate time frame, the BS will transmit these packets to RS group g through the relay link transmission. These packets received by RS group g are also buffered in the packet queue corresponding to MS 1 until the access link transmission between RS group g and MS 1 is scheduled in a following frame.

To realize this two-hop relaying operation, the scheduling procedure should be performed by the BS at the beginning of each frame. Specifically, the scheduling procedure consists of two parts: 1) *relay link scheduling* and 2) *access link scheduling*. In the relay link scheduling, the BS selects an RS group. The relay link between the BS and this RS group will be activated during the access zone to transmit packets destined to the MSs residing in this RS group. These packets will be queued in the RS group for future relaying. On the other hand, in access-link scheduling, the BS selects an appropriate activation set of RS groups. The access links between these RS groups and their respective resident MSs will be activated during the relay zone for packet relaying. Note that, since the BS has only one radio transmitter, only one relay link between the BS and a particular RS group can be activated at a time in the access zone of each time frame. However, based on the concept of spatial reuse, more than one access link transmission can simultaneously be activated in the relay zone.

To resolve the scheduling problem for RS-grouping-enabled IEEE 802.16j MR networks, scheduling policies should be employed to decide suitable relay link transmissions in access zones and access link transmissions in relay zones. Different scheduling policies may be designed based on different criteria to meet different system performance requirements.

B. Radio Resource-Scheduling Policies

1) System Model: Consider again the single-BS IEEE 802.16j MR network in Fig. 5. Using our RS grouping algorithm, the RSs in this network are partitioned into a set Γ of RS groups, and the RS groups are classified into a set Ω of activation sets. The RS groups within each activation set can simultaneously be activated.

For both the relay link and access link transmissions in the considered IEEE 802.16j MR network, the Rayleigh-fading channel model is assumed to be applied in the physical layer. Letting $src(l)$ be the transmitter and $dst(l)$ be the receiver of a radio link l, we can compute the transmission rate over the radio link l as

$$R_l = B \cdot \log_2 \left(1 + \left|h_{dst(l)}^{src(l)}\right|^2 \cdot PW_{src(l)}/\sigma_N^2\right). \quad (1)$$

In (1), B denotes the bandwidth (in hertz) of the MR network spectrum, and $PW_{src(l)}$ and σ_N^2 represent the transmission power of $src(l)$ and the variance of signal noise, respectively. Moreover, $h_{dst(l)}^{src(l)}$ in (1) is a circularly symmetric complex Gaussian random variable. See [28] for a detailed explanation of the derivation of (1).

During the relay link transmission, the BS adopts the minimum transmission rate among the relay links between the BS and the selected RS group as the real transmission rate. This is due to the fact that the RSs within the selected RS group must correctly receive and decode the data from the BS at the same time and frequency. In this case, the effective transmission rate is dominated by the relay link with the minimum rate. Thus, the transmission rate of the relay link between the BS and an RS group g can be computed based on (1) as

$$R_{\mathrm{BS},g} = B \cdot \log_2 \left(1 + \min_{\forall r \in g} \left|h_r^{\mathrm{BS}}\right|^2 \cdot PW_{\mathrm{BS}}/\sigma_N^2\right). \quad (2)$$

For the access link transmission, although the RS members in the same RS group should serve a particular MS at the same time, this signal-combining cooperative transmission with multiple sources is difficult to implement. Therefore, we adopt a compromise cooperation scheme called *selection relaying* [15] to realize the access link transmission, whereby only the RS member with the highest signal quality is selected for data transmission toward a resident MS. Therefore, the transmission rate between the RS group g and a resident MS j can be derived similar to (2) as

$$R_{g,j} = B \cdot \log_2 \left(1 + \max_{\forall r \in g} \left|h_j^r\right|^2 \cdot PW_r/\sigma_N^2\right). \quad (3)$$

In the two proposed scheduling policies that will be described later, the network queue state is a common parameter utilized to make the scheduling decision. We denote $X_j^{\mathrm{BS}}(t)$ as the queue length of the queue corresponding to MS j in the BS at the start of time frame t. In the RS group g of MS j, every RS member will also maintain a queue for MS j. However, as previously mentioned, all the RS members serve MS j at the same time. Thus, from MS j's viewpoint, the queues maintained in the RSs for MS j can be regarded as a logical single queue. We denote $X_j^g(t)$ as the length of the logical queue for MS j in RS group g at the start of time frame t.

Note that the RS buffer synchronization for selection relaying can easily be achieved by utilizing the standard IEEE 802.16j ARQ mechanism and by arranging a new multicast information element in the DL-MAP. More specifically, when the BS receives an ACK from an MS indicating the sequence number of the latest correctly decoded packet, it will, in the next time frame, announce an additional DL-MAP multicast information element containing this sequence number to the corresponding RS group. The obsolete packets at each individual RS buffer can then be removed accordingly.

2) General Scheduling Procedure of the RS-Grouping-Enabled IEEE 802.16j MR Network: Recall that the scheduling procedure should make the decision of the relay link and access link transmissions of a frame in the relay link scheduling and access link scheduling, respectively. Although the selection criteria of these two parts are actually the same under the given scheduling policy, some detailed processes are different due to the special characteristics of RS grouping, such as minimum-transmission rate constraints and selection-relaying cooperative transmissions. Therefore, we individually detail the procedures of these two scheduling parts as follows:

Part 1: Relay link scheduling. In this part, the objective of the scheduler is to activate the most favorable relay link between the BS and an RS group in Γ. For comparison, we claim that the major task of a scheduling policy is to assign a weighted index $D_g^{\mathrm{RL}}(t)$ for each RS group g to represent the priority of RS group g to be selected in the relay link scheduling of time frame t. Note that the implementation of $D_g^{\mathrm{RL}}(t)$ is dependent on the concerned performance metrics (e.g., throughput or delay time) of the scheduling policy. To make the scheduling decision in a channel-aware manner, the scheduling policy may require the transmission rate $R_{\mathrm{BS},g}$ to decide the $D_g^{\mathrm{RL}}(t)$ of RS group g. Two different approaches are given later in this section. After both the transmission rate and the weighted index computations of each RS group, the final target RS group $\hat{g}(t)$ for time frame t is selected as follows:

$$\hat{g}(t) = \arg\max_{g \in \Gamma} \left\{D_g^{\mathrm{RL}}(t)\right\}. \quad (4)$$

In other words, the RS group with the maximum weighted index will be selected as the target RS group. During the target RS group selection, the set $\hat{I}_g(t)$ of MSs whose data bursts will be delivered to the chosen RS group $\hat{g}(t)$ should also be identified. This MS identification is based on the underlying performance metrics as well. Following the scheduling result, the data bursts belonging to these MSs in $\hat{I}_g(t)$ are scheduled in the access zone of the time frame t for transmission to $\hat{g}(t)$. All RS members of $\hat{g}(t)$ should receive and buffer the data for the purpose of second-hop transmission.

Part 2: Access link scheduling. To achieve simultaneous access link transmissions in each time frame, the scheduler is responsible for selecting an appropriate activation set and the corresponding target MSs in access link scheduling. Likewise, a scheduling policy is also needed to assign a weighted index $D_g^{\mathrm{AL}}(t)$ for each RS group g to denote the priority of RS group g to be selected in the access link scheduling of time frame t. The transmission rate $R_{g,j}$ may also be required to decide the $D_g^{\mathrm{AL}}(t)$ of RS group g by the scheduling policy. Since the activation set assignment is provided *a priori*, the activation set $\hat{\phi}(t)$ with the maximum total RS group weighted indices can be selected from Ω for transmission at time frame t, i.e.,

$$\hat{\phi}(t) = \arg\max_{\phi \in \Omega} \left\{ \sum_{g \in \phi} D_g^{\mathrm{AL}}(t) \right\}. \quad (5)$$

Regarding the choice of MSs to be served in the access link scheduling, we denote $\hat{j}_g(t)$ as the MS residing in RS group $g \in \hat{\phi}(t)$ whose packets will be transmitted over the selected cooperative access link at time frame t. In the following two proposed scheduling policies, the respective $\hat{j}_g(t)$ are provided in (8) and (14).

3) TF Scheduling Policy: Different scheduling policies may be applied in the scheduling procedure to provide the different definitions of the weighted indices $D_g^{\mathrm{RL}}(t)$ and $D_g^{\mathrm{AL}}(t)$ and the MS selections $\hat{I}_g(t)$ and $\hat{j}_g(t)$ based on their respective performance criteria. Similar to the throughput-optimal scheduling policy mentioned in [11], we introduce a throughput-first (TF) scheduling policy, which is compatible with the scheduling procedure described in Section V-B2. The objective of the TF policy is to achieve the maximum system throughput for our considered IEEE 802.16j MR network.

Consider the scheduling procedure using the TF scheduling policy. Let Λ_g be the set of MSs residing in the same group g. In the relay link scheduling of time frame t, the weighted index $D_g^{\mathrm{RL}}(t)$ of a group $g \in \Gamma$ is defined as

$$D_g^{\mathrm{RL}}(t) = R_{\mathrm{BS},g} \sum_{j \in \Lambda_g} X_j^{\mathrm{BS}}(t) \quad (6)$$

where $R_{\mathrm{BS},g}$ is computed from (2). In access link scheduling, the weighted index $D_g^{\mathrm{AL}}(t)$ is defined as

$$D_g^{\mathrm{AL}}(t) = \max_{j \in \Lambda_g} \left[R_{g,j} \cdot X_j^g(t) \right] \quad (7)$$

where $R_{g,j}$ is computed from (3). By (6) and (7), the transmission links with larger queue length and higher transmission rate are desired to be served. For the demonstration purpose, the simple round-robin approach is applied to determine the $\hat{I}_g(t)$ set during the relay link scheduling. On the other hand, during access link scheduling, the MS selection has already been done in the computation of $D_g^{\mathrm{AL}}(t)$. That is, if a group $g \in \hat{\phi}(t)$ is selected to be active at time frame t, the MS whose packets will be transmitted by the group g is

$$\hat{j}_g(t) = \arg\max_{j \in \Lambda_g} \left[R_{g,j} \cdot X_j^g(t) \right]. \quad (8)$$

4) DF Scheduling Policy: Since delay-sensitive applications (e.g., Voice over Internet Protocol call) are widely used, a delay-oriented scheduling policy for IEEE 802.16j MR networks is desired. With compatibility to the scheduling procedure described in Section V-B2, we propose a delay-first (DF) scheduling policy, which is aimed at minimizing the average packet delay time. The concept of the DF scheduling policy is to first serve the user queue with the maximum predicted mean packet waiting time.

We assume that there is an observation time window with length T_w. The average packet arrival rate $\lambda_j^i(t)$ of MS j on a transmitter i (the BS or an RS group) of the T_w frames before time frame t can be estimated with an iterative approach as follows:

$$\lambda_j^i(t) = \left(1 - \frac{1}{T_w}\right) \lambda_j^i(t-1) + \frac{1}{T_w} a_j^i(t) \quad (9)$$

where $a_j^i(t)$ denotes the packet arrival rate of MS j on the transmitter i in time frame t. Note that the $\lambda_j^i(t)$ in (9) is updated at the end of each frame transmission on all transmitters. Similarly, the mean queue length $\bar{X}_j^i(t)$ of MS j on a transmitter i of the T_w frames before time frame t can be estimated as

$$\bar{X}_j^i(t) = \left(1 - \frac{1}{T_w}\right) \bar{X}_j^i(t-1) + \frac{1}{T_w} X_j^i(t). \quad (10)$$

Therefore, by Little's formula [18], the predicted mean packet waiting time of MS j on the transmitter i in time frame t can be written as

$$W_j^i(t) = \frac{\bar{X}_j^i(t)}{\lambda_j^i(t)}. \quad (11)$$

Let $W_j^{\mathrm{BS}}(t)$ and $W_j^g(t)$ be the predicted mean waiting times for the packets of MS j at the BS and an RS group $g \in \Gamma$, respectively. In relay link scheduling, the weighted index $D_g^{\mathrm{RL}}(t)$ of RS group g is defined as

$$D_g^{\mathrm{RL}}(t) = \sum_{j \in \Lambda_g} W_j^{\mathrm{BS}}(t). \quad (12)$$

Moreover, the weighted index $D_g^{\mathrm{AL}}(t)$ in access link scheduling is formulated as

$$D_g^{\mathrm{AL}}(t) = \max_{j \in \Lambda_g} W_j^g(t). \quad (13)$$

Similarly, for demonstration purposes, the $\hat{I}_g(t)$ set for relay link transmission is also determined in a round-robin manner. Finally, if a group $g \in \hat{\phi}(t)$ is selected to be active in time frame t by access link scheduling, the MS whose packets will be transmitted by the group g is

$$\hat{j}_g(t) = \arg\max_{j \in \Lambda_g} W_j^g(t). \quad (14)$$

VI. Simulation Results and Discussion

In this section, we present the performance evaluations of the RS grouping algorithm and centralized DL scheduling policies. We consider an IEEE 802.16j MR network with a 10-MHz system bandwidth and a 2.5-GHz carrier frequency. Note that only the DL transmission is discussed in this paper; therefore, the MAC frame structure in our simulation experiments is simplified to comprise only the DL subframe in each time frame. The MAC frame length is set to 10 ms. Since we only focus on the performance of resource allocation, the controlling signal overheads are not considered in the simulation, and the durations of the access zone and the relay zone within each

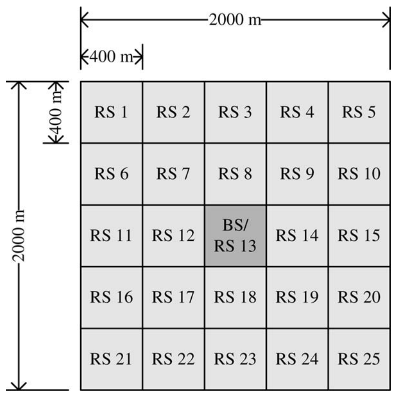

Fig. 6. Simulation topology. Single BS-cell system with a 25-RS-cell grid layout.

MAC frame are set to 0.5 and 9.5 ms, respectively. Moreover, we assume that the relay link (BS-to-RS) is operated under the LOS channel condition; thus, the target throughput rate R_{TH} is assumed to be 60 Mb/s. On the other hand, for the access link (RS-to-MS) transmission, the NLOS channel condition with $R_{\mathrm{TH}} = 40$ Mb/s is considered.

To reduce our simulation complexity, the square-based cellular topology shown in Fig. 6 is adopted in our simulation environment. We consider a single square-shaped BS-cell topology where the BS is at the center, and 25 square-shaped RS cells are regularly organized to form a 5 × 5 grid layout. The widths of the BS cell and RS cell are 2000 and 400 m, respectively. We also consider that a number of MSs are uniformly distributed in the BS cell. Note that we adopt the uniform Random-Walk mobility model [16] to realize the MS movement behavior. Based on the handoff-rate information from this mobility model, our RS grouping algorithm generates the RS grouping layouts with determined activation set assignments, as shown in Fig. 7 (among the 25 RS grouping layouts, only the layouts for the preferred group sizes 1–12 are demonstrated). The packet arrivals of each user at the BS follow a Poisson process with the mean interarrival time setting to 100 ms, and the packet length is assumed to be exponentially distributed. Moreover, we equally divide all the users into five classes, which are offered with different DL traffic loads. Except for the experiment of Fig. 9 (which investigates the effects of traffic loads and RS group sizes on system stability), the mean packet length of class i users is set to $i \times 1000$ bits (e.g., the mean packet lengths of class-1 and class-5 users are 1000 and 5000 bits, respectively).

A. *Effects of RS Group Sizes on Handoff Frequency*

Fig. 8 individually evaluates the handoff frequency of each grouping in Fig. 7 under different user densities (specifically, 100, 500, and 1000 users in our experiments). The results indicate that, using the greedy grouping policy of our RS grouping algorithm, the handoff frequencies of the groupings gradually decline, regardless of the user density, as the group size increases. The main reason is that larger group sizes generally lead to lower handoff probabilities. It could be concluded that it is reasonable to choose larger group sizes if higher user mobility speeds are observed.

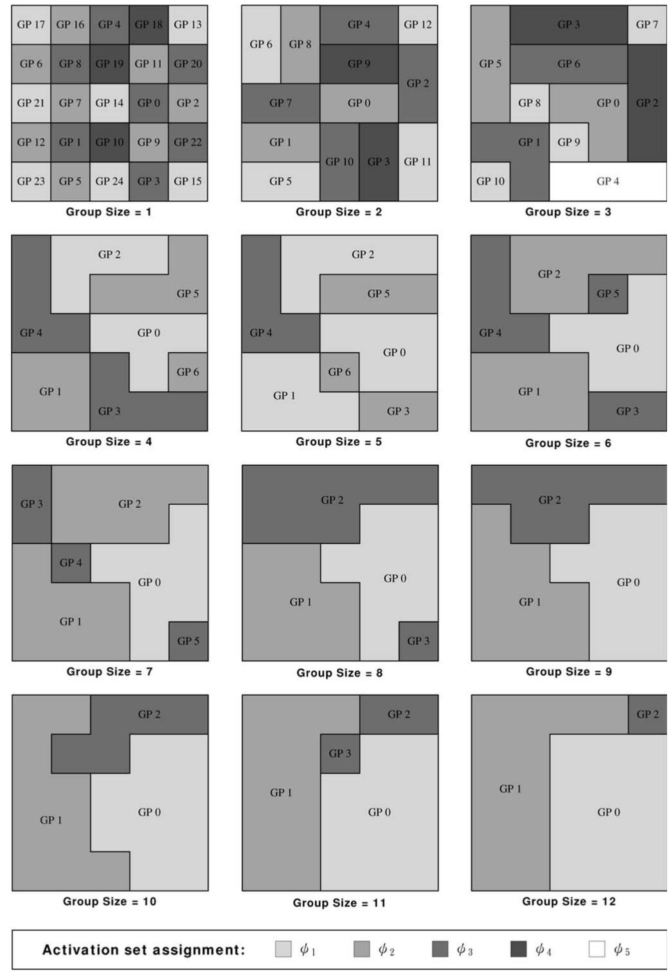

Fig. 7. Generated grouping results with activation set assignments.

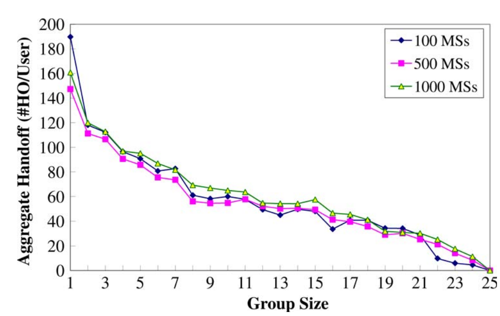

Fig. 8. Handoff comparison between different group sizes and different user densities.

B. *Effects of RS Group Sizes on System Stability*

Fig. 9 shows the mean packet delay under the grouping scenarios with group sizes 2, 6, and 10, as shown in Fig. 7. The delay performance is expressed as a function of the average packet traffic load to 100 MSs. For this experiment, we apply the TF policy during the scheduling procedure. Under the grouping scenario with the largest group size (i.e., group size of 10), the system becomes unstable (i.e., the users' packet delays go to infinity) when the average packet traffic load is larger than 1000 kb/s. Nevertheless, decreasing the group size can gradually extend the stable region. For example, in the grouping scenarios with group sizes 6 and 2, the stable regions

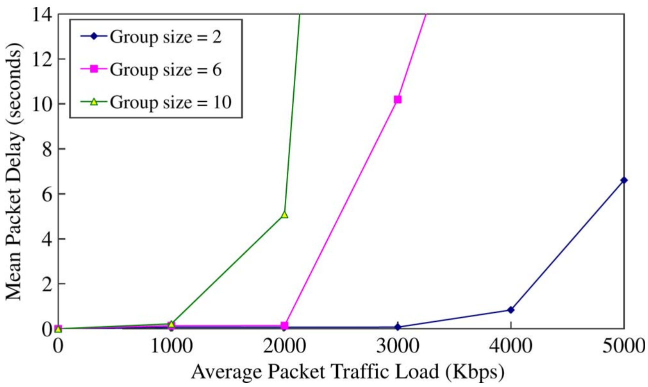

Fig. 9. Mean packet delay versus average packet traffic load under groupings with group sizes 2, 6, and 10.

are enhanced to 2000 and 4000 kb/s, respectively. These results are primarily due to the spatial diversity gain discrepancies under different grouping scenarios. Specifically, groupings with smaller group sizes may have more opportunities to perform simultaneous transmissions, which means the system capacities can thus be reasonably enlarged. Note that smaller group sizes may, however, lead to more handoff events.

C. Performance Analysis of Different Scheduling Policies

Based on the groupings shown in Fig. 7, we estimate the throughput and delay performances of the TF, DF, and Round-Robin centralized scheduling policies as functions of the group size. Figs. 10 and 11 present the average throughput and delay of the network under the situations with fixed and mobile users, respectively. Note that the average throughput is estimated as the average amount of packets actually received by the users within the simulation time. Moreover, the packet delay indicates the time difference between when a packet arrives at the BS and when it is received by the corresponding user.

Figs. 10(a) and 11(a) show the average network throughput under the three scheduling policies. For the case where users are stationary, the TF and DF policies achieve similar throughput performance. However, for the case with user mobility, the average throughput under the DF policy is higher than those under the other two scheduling policies. This phenomenon implies that the mobility behavior of the MSs is a major factor affecting the throughput performance under the TF policy, whereas the DF policy can provide relatively stable throughput, regardless of the user-mobility effect. Specifically, since the DF policy is able to balance the queue length of each user while the TF policy only prioritizes the user with the longest queue length, the average buffered packet amounts for all users under the DF policy are relatively small. Therefore, the number of dropped packets during the handoffs in DF can reasonably be reduced, and the throughput can further be improved. The Round-Robin policy inefficiently performs in both cases since its scheduling decisions are sequentially made, despite the system state information. In addition, we also observe that, compared with groupings with larger group sizes, groupings with smaller group sizes result in better throughput performance for the case of fixed users. However, when user mobility is considered, the throughput value increases as the group size increases. This is because, when users are stationary, the throughput performance is mainly dominated by the spatial diversity gain. Under such circumstances, smaller RS group sizes lead to higher throughput values. On the other hand, when the user mobility speed is high enough, the larger packet loss rates due to more frequent handoffs of smaller group sizes have more significant impact on the throughput performance. Consequently, larger RS group sizes are more advantageous.

Figs. 10(b) and 11(b) show the packet delay performance for the two scenarios without and with user mobility. We notice that groupings with larger group sizes result in higher average packet delay. This is owing to the lower spectrum reuse of larger group sizes, and the packets in the system may suffer from longer queueing delay. Moreover, as expected, the DF policy can provide the lowest average delay for all group size cases, compared with the TF and Round-Robin policies. The TF policy causes higher average delay since it considers only the queue length states but no packet waiting time information. The Round-Robin policy, which depends on neither the queue length nor the waiting time information, incurs the worst average delay performance through almost all group size cases.

D. Effects of RS Grouping Patterns

In Figs. 10 and 11, we also observe that the performance under the case of group size 3 is worse than those under the other cases. Two primary reasons for this phenomenon are given here. First, the activation sets used in this case are more than those used in all the other cases (see Fig. 7). As discussed in Section IV-C, the more activation sets there are, the less the spatial diversity gain. Therefore, poorer throughput and delay performance results are expected for group size 3. Second, since the group size in this case is small, the packet losses from handoff events would also influence the throughput performance. Under the joint impacts of low spatial diversity gain and high packet loss rate, the worst performances are observed for the case of group size 3. This special case implies that, if the RS grouping patterns cannot appropriately be determined, the system performance would significantly be affected. The simulation results shown in Figs. 10 and 11 demonstrate that our proposed RS grouping algorithm can derive suitable RS grouping patterns in most cases.

E. Fairness Analysis

The fairness properties of the TF, DF, and Round-Robin scheduling policies are shown in Fig. 12, where the 100 users are classified into five classes. The scenario with group size 20 is considered in the experiments. We use the average delay for the interclass fairness comparison. On the other hand, for the intraclass fairness comparison, we use the per-user packet delay to compute the Jain Fairness Index [24] as follows:

$$\text{Fairness Index} = \frac{\left(\sum_1^n x_i\right)^2}{n\sum_1^n x_i^2} \tag{15}$$

where x_i is the average packet delay of user i. A scheduling policy achieves the optimal fairness if this index is 1 and is completely unfair if it is $1/n$.

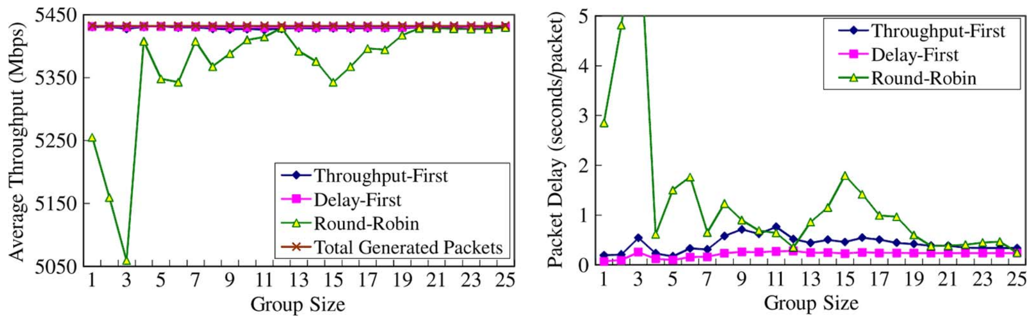

Fig. 10. Performance comparison between different group sizes in terms of (a) throughput and (b) delay without user mobility.

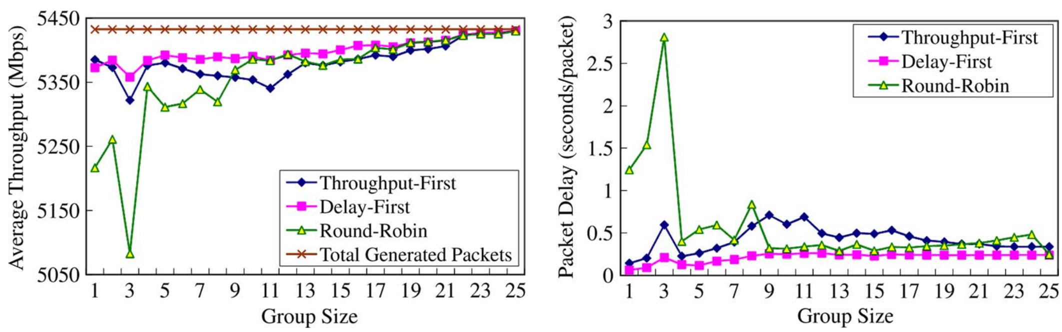

Fig. 11. Performance comparison between different group sizes in terms of (a) throughput and (b) delay with user mobility.

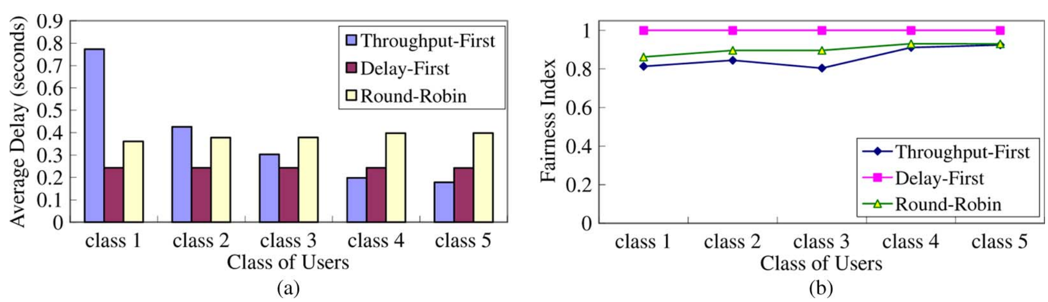

Fig. 12. Fairness comparison in terms of packet delay.

It can be observed that fairness among users is not satisfied under the TF scheduling policy. Since the high-load users have high service priority, the average delays of these users are significantly low. However, the users who have lighter offered loads tend to suffer from higher packet delay. On the other hand, both the DF and Round-Robin scheduling policies balance each user's average delay, which implies that fairness among users (whether interclass or intraclass) is guaranteed in terms of the average delay under these two policies. Moreover, the DF scheduling policy provides lower average delay than the Round-Robin policy for all the five classes of users, because the maximum mean waiting time of all users' queues is minimized under the DF policy. These observations lead to the conclusion that our DF scheduling policy can provide not only the minimization of the average delay but the fairness property among different traffic-load users as well.

VII. Related Work

In the literature, registration area (RA) grouping schemes and radio resource-scheduling policies have been intensively investigated for mobile telecommunications networks and for MR networks, respectively. However, we note that, in these studies, either scheduling issues were not addressed in the RA grouping schemes or grouping concepts were not considered in the MR scheduling policies. Grouping mechanisms and scheduling algorithms have never been jointly discussed. Here, we separately summarize some representative RA grouping schemes and MR scheduling policies.

1) *RA grouping schemes:* In mobile telecommunications networks, several cells constitute an RA. How to group cells to form RAs is an important cellular network-planning issue. In [12], Xie *et al.* first proposed a per-user adaptive RA scheme. By taking paging rates, location-update rates, and the respective costs into account, the proposed scheme decides the appropriate size of an RA to minimize the overall cost. Then, Castelluccia addressed the same grouping problem in [2] by considering a more general network configuration that supports overlapping RAs. Grouping users into different velocity classes according to their velocities upon location update instants, the scheme in [10] reduces the paging cost by properly decreasing RA sizes. To identify the optimal RA size and

cope with the location-tracking problem, a dual-group system was proposed in [3]. The dual-group method enhances the utilization of network resources, because it eliminates unnecessary location updates due to its fewer group boundary passes. Dynamically adaptive grouping schemes were also discussed in [8] and [9]. In [8], the proposed scheme dynamically adjusts overlapping RAs according to user mobility and call patterns. In this paper, the effects of cell inclusions/exclusions into/from an RA were analytically characterized. By periodically determining the proper number of cells within an RA, the proposed scheme greatly reduces the average signaling cost and database access load on location registers. In [9], the authors further demonstrated how the grouping of an RA affects the number of subsequent registrations in the overlapping RA environments.

2) *MR scheduling policies:* Various scheduling policies have been proposed to achieve different performance objectives, such as throughput maximization [11], [22], [29], delay minimization [23], proportional fairness [20], [26], [27], and quality-of-service guarantee [4], [5], [17], [21]. According to specific objectives, the resource scheduling problem can be formulated to different optimization problems. For instance, the scheduling problem was modeled as a link-selection problem in [11] to maximize the signal-to-interference-plus-noise ratio of the links and to prevent all the user queues from exploding, a boundary selection problem in [26] to balance the traffic loads among relays, a graph multicoloring problem in [29] to maximize the sizes of nonadjacent-relay sets, a user-required-rate problem in [21] to satisfy each user's required rate, a network flow problem in [23] to select an appropriate transmission order, and an interference-minimization problem in [17] to find out a path with less interference for each connection. To resolve these problems in polynomial time, heuristic scheduling policies were proposed in the respective research.

VIII. Conclusion

The IEEE 802.16j MR standard has been developed to provide performance enhancement to the existing IEEE 802.16e network. However, issues such as frequent handoffs and low spectrum utilization, which were not encountered in IEEE 802.16e, may occur in IEEE 802.16j. The RS grouping is one optional mechanism in the IEEE 802.16j MR standard to overcome these problems. This paper has investigated the RS grouping performance enhancement in terms of throughput and handoff frequency. An RS grouping algorithm was designed by utilizing a greedy grouping policy: RS pairs with higher handoff rates will have higher priority to be selected. The simulation results have shown that the handoff frequency of the considered MR network can significantly be reduced, and suitable RS grouping patterns with determined activation set assignments can be derived using our grouping algorithm. In addition, we have proposed the TF and DF centralized scheduling policies to maximize the system throughput and to minimize the average packet delay, respectively. By integrating our RS grouping algorithm and centralized-scheduling algorithms, the simulation results have indicated that, for the case of fixed users, groupings with smaller group sizes can result in better throughput performance. However, when user mobility is considered, the throughput value increases as the group size increases. Furthermore, we have also shown that the DF policy cannot only minimize the average packet delay but can also provide fairness among different traffic-load users.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments, which have significantly enhanced the quality of this paper.

References

[1] *Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*, IEEE Std. 802.16e-2005, Feb. 2006.

[2] C. Castelluccia, "Extending mobile IP with adaptive individual paging: A performance analysis," in *Proc. IEEE ISCC*, 2000, pp. 113–118.

[3] C.-M. Weng and P.-W. Huang, "Modified group method for mobility management," *Comput. Commun.*, vol. 23, no. 2, pp. 115–122, Jan. 2000.

[4] C.-Y. Hong and A.-C. Pang, "Link scheduling with QoS guarantee for wireless relay networks," in *Proc. IEEE 28th INFOCOM*, Apr. 2009, pp. 2806–2810.

[5] D. Ghosh, A. Gupta, and P. Mohapatra, "Admission control and interference-aware scheduling in multi-hop WiMAX networks," in *Proc. IEEE MASS*, Oct. 2007, pp. 1–9.

[6] D. J. A. Welsh and M. B. Powell, "An upper bound for the chromatic number of a graph and its application to timetabling problems," *Comput. J.*, vol. 10, no. 1, pp. 85–86, 1967.

[7] D. Schultz, R. Pabst, and T. Irnich, "Multi-hop based radio network deployment for efficient broadband radio coverage," in *Proc. WPMC*, Yokosuka, Japan, 2003, pp. 377–381.

[8] G. Varsamopoulos and S. K. S. Gupta, "Dynamically adapting registration areas to user mobility and call patterns for efficient location management in PCS networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 5, pp. 837–850, Oct. 2004.

[9] G. Varsamopoulos and S. K. S. Gupta, "Optimal offline and online registration techniques for location management with overlapping registration areas," *IEEE Trans. Mobile Comput.*, vol. 4, no. 5, pp. 474–488, Sep./Oct. 2005.

[10] G. Wan and E. Lin, "A dynamic paging scheme for wireless communication systems," in *Proc. ACM/IEEE MobiCom*, 1997, pp. 195–203.

[11] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2318–2328, Sep. 2005.

[12] H. Xie, S. Tabbane, and D. J. Goodman, "Dynamic location area management and performance analysis," in *Proc. IEEE Veh. Technol. Conf.*, May 1993, pp. 536–539.

[13] *Ranging Process for IEEE 802.16j*, IEEE 802.16j-06/172, Nov. 2006.

[14] *Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification*, IEEE Std. 802.16j-2009, Dec. 2009.

[15] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[16] J. Nuevo, Mobility Generator for NS-22002. [Online]. Available: http://www.math.keio.ac.jp/matumoto/emt.html

[17] J. Tang, G. Xue, and C. Chandler, "Interference-aware routing and bandwidth allocation for QoS provisioning in multihop wireless networks," *Wireless Commun. Mobile Comput. J.*, vol. 5, no. 8, pp. 933–943, Dec. 2005.

[18] L. Kleinrock, *Queueing Systems*. Hoboken, NJ: Wiley, 1975.

[19] L. Kleinrock and J. Silvester, "Spatial reuse in multihop packet radio networks," *Proc. IEEE*, vol. 75, no. 1, pp. 156–167, Jan. 1987.

[20] L. Wang, Y. Ji, and F. Liu, "A novel centralized resource scheduling scheme in OFDMA-based two-hop relay-enhanced cellular systems," in *Proc. IEEE WIMOB*, Oct. 2008, pp. 113–118.

[21] M. K. Awad and X. Shen, "OFDMA based two-hop cooperative relay network resources allocation," in *Proc. IEEE ICC*, May 2008, pp. 4414–4418.

[22] M. Yang and P. H. J. Chong, "Time slot allocation schemes for multi-hop TDD-CDMA cellular system," in *Proc. IEEE WCNC*, Mar. 2007, pp. 3099–3104.

[23] P. Djukic and S. Valaee, "Link scheduling for minimum delay in spatial re-use TDMA," in *Proc. IEEE 26th INFOCOM*, May 2007, pp. 28–36.

[24] R. Jain, W. Hawe, and D. Chiu, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corp., Maynard, MA, Tech. Rep. DEC-TR-301, Sep. 1984.

[25] R. Pabst, B. H. Walke, and D. C. Schultz, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sep. 2004.

[26] T.-W. Kim, T.-Y. Min, and C.-G. Kang, "Opportunistic packet scheduling algorithm for load balancing in a multi-hop relay-enhanced cellular OFDMA-TDD system," in *Proc. 14th APCC*, Oct. 2008, pp. 1–5.

[27] T. Wu, G. Li, Y. Wang, J. Huang, X. Yu, and H. Tian, "Fairness-oriented scheduling with equilibrium for multihop relaying networks based on OFDMA," in *Proc. IEEE 68th VTC—Fall*, Sep. 2008, pp. 1–5.

[28] W.-C. Kan and S.-R. Yang, The Supplements to "Handoff minimization through a relay station grouping algorithm with efficient radio resource scheduling policies for IEEE 802.16j multihop relay networks," Nat. Tsing Hua Univ., Hsinchu, Taiwan. [Online]. Available: http://www.cs.nthu.edu.tw/~sryang/submission/TechHMRSGA.pdf

[29] X. Guo, W. Ma, Z. Guo, X. Shen, and Z. Hou, "Adaptive resource reuse scheduling for multihop relay wireless network based on multicoloring," *IEEE Commun. Lett.*, vol. 12, no. 3, pp. 176–178, Mar. 2008.

Shun-Ren Yang (M'08) received the B.S. and M.Sc. degrees in computer science and information engineering and the Ph.D. degree from National Chiao Tung University, Hsinchu, Taiwan, in 1998, 1999, and 2004, respectively.

From April 1, 2004, to July 31, 2004, he was a Research Assistant with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong. Since August 2004, he has been with the Department of Computer Science and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, where he is currently an Assistant Professor. His current research interests include the design and analysis of personal communications service networks, computer telephony integration, mobile computing, and performance modeling.

Prof. Yang is a member of the IEEE Computer and IEEE Communications Societies.

Chien-Chi Kao (S'09) received the B.S. degree in computer science and information engineering from National Chung Cheng University, Chiayi, Taiwan, in 2006 and the M.S.Com. degree from National Tsing Hua University, Hsinchu, Taiwan, in 2008. He is currently working toward the Ph.D. degree with the Department of Computer Science, National Tsing Hua University.

In 2006, he joined the Wireless and Mobile Network Laboratory, Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University. His current research interests include the design and analysis of mobile telecommunications networks and performance modeling.

Wai-Chi Kan received the B.S. degree in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2006 and the M.S.Cs. degree from National Tsing Hua University, Hsinchu, in 2008.

In 2006, he joined the Wireless and Mobile Network Laboratory, Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University. Since 2008, he has been with Synology Inc., Taipei, Taiwan. His current research interests include the design and analysis of mobile telecommunications networks and performance modeling.

Tzung-Chin Shih received the B.S. degree in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2008. He is currently working toward the Master's degree with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu.

In 2008, he joined the Wireless and Mobile Network Laboratory, Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University. His current research interests include the design and analysis of mobile telecommunications networks and performance modeling.